# wwPDB X-ray Structure Validation Summary Report (i)

Mar 4, 2024 – 12:40 AM EST

| | | |
|---|---|---|
| PDB ID | : | 4HP7 |
| Title | : | Trioxacarcin D517 as a product of guanine robbery from d(AACCGGTT) |
| Authors | : | Smaltz, D.J.; Magauer, T.; Proepper, K.; Dittrich, B.; Myers, A.G. |
| Deposited on | : | 2012-10-23 |
| Resolution | : | 1.09 Å(reported) |

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

---

The following versions of software and data (see references (i)) were used in the production of this report:

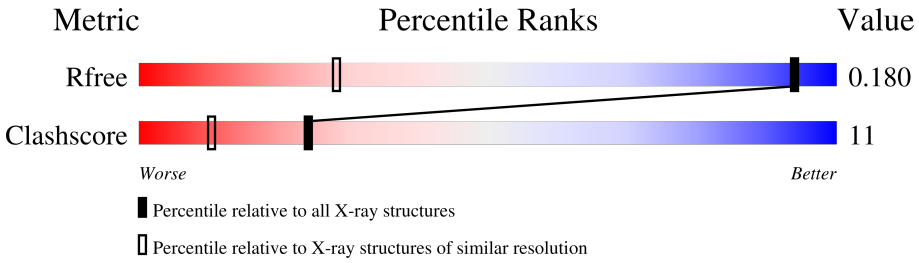
| | | |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 1.8.5 (274361), CSD as541be (2020) |
| Xtriage (Phenix) | : | 1.13 |
| EDS | : | 2.36 |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20191225.v01 (using entries in the PDB archive December 25th 2019) |
| Refmac | : | 5.8.0158 |
| CCP4 | : | 7.0.044 (Gargrove) |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.36 |

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.09 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



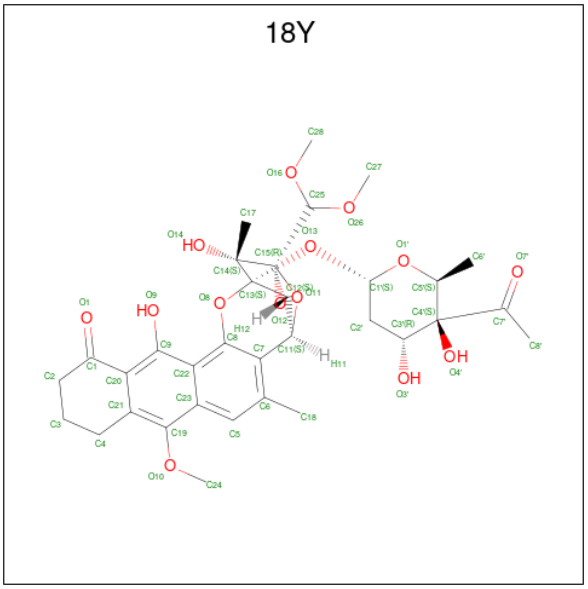

| Metric | Whole archive (#Entries) | Similar resolution (#Entries, resolution range(Å)) |
|---|---|---|
| $R_{free}$ | 130704 | 1619 (1.14-1.06) |
| Clashscore | 141614 | 1671 (1.14-1.06) |

## 2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 451 atoms, of which 192 are hydrogens and 0 are deuteriums.

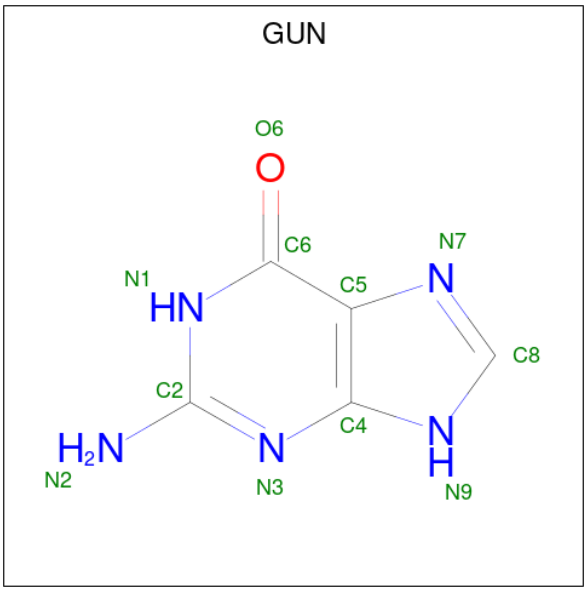
In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is Trioxacarcin A analogue, bound form (three-letter code: 18Y) (formula: $C_{33}H_{40}O_{14}$).



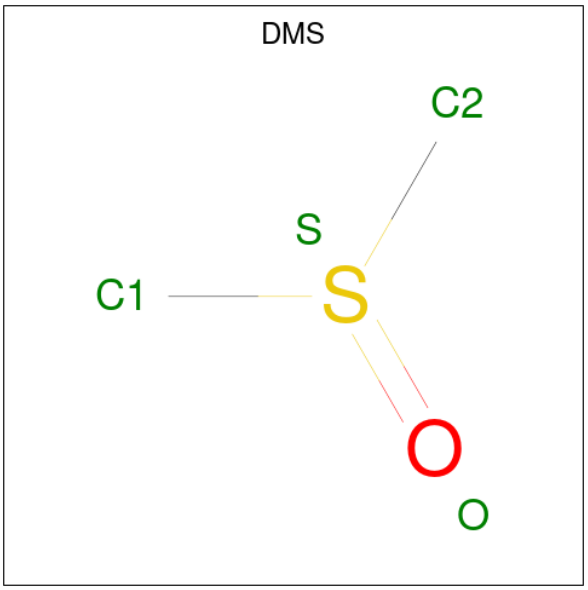

| Mol | Chain | Residues | Atoms | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|
| 1 | A | 1 | Total | C | H | O | 0 | 1 |
| | | | 90 | 34 | 41 | 15 | | |
| 1 | B | 1 | Total | C | H | O | 0 | 0 |
| | | | 84 | 33 | 37 | 14 | | |
| 1 | C | 1 | Total | C | H | O | 0 | 0 |
| | | | 84 | 33 | 37 | 14 | | |
| 1 | D | 1 | Total | C | H | O | 0 | 1 |
| | | | 86 | 33 | 37 | 16 | | |

- Molecule 2 is GUANINE (three-letter code: GUN) (formula: $C_5H_5N_5O$).




| Mol | Chain | Residues | Atoms | ZeroOcc | AltConf |
|---|---|---|---|---|---|
| 2 | A | 1 | Total C H N O<br>15 5 4 5 1 | 0 | 1 |
| 2 | B | 1 | Total C H N O<br>15 5 4 5 1 | 0 | 0 |
| 2 | C | 1 | Total C H N O<br>15 5 4 5 1 | 0 | 0 |
| 2 | D | 1 | Total C H N O<br>15 5 4 5 1 | 0 | 0 |

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: $C_2H_6OS$).



| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|---|
| 3 | C | 1 | Total | C | H | O | S | 0 | 0 |
| | | | 10 | 2 | 6 | 1 | 1 | | |
| 3 | C | 1 | Total | C | H | O | S | 0 | 0 |
| | | | 10 | 2 | 6 | 1 | 1 | | |
| 3 | D | 1 | Total | C | H | O | S | 0 | 1 |
| | | | 20 | 4 | 12 | 2 | 2 | | |

- Molecule 4 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| 4 | D | 7 | Total | O | 0 | 0 |
| | | | 7 | 7 | | |

# 3 Residue-property plots (i)

There is no protein, DNA or RNA chain in this entry to show sequence plots.

# 4 Data and refinement statistics (i)

| Property | Value | Source |
|---|---|---|
| Space group | P 1 21 1 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 24.09Å 9.83Å 35.09Å<br>90.00° 92.32° 90.00° | Depositor |
| Resolution (Å) | 40.00 – 1.09<br>35.06 – 1.09 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 96.6 (40.00-1.09)<br>97.0 (35.06-1.09) | Depositor<br>EDS |
| $R_{merge}$ | 0.10 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 0.20 (at 1.09Å) | Xtriage |
| Refinement program | SHELXL 2012 | Depositor |
| R, $R_{free}$ | 0.153 , 0.216<br>0.176 , 0.180 | Depositor<br>DCC |
| $R_{free}$ test set | 4256 reflections (10.16%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 8.0 | Xtriage |
| Anisotropy | 0.102 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.41 , 104.4 | EDS |
| L-test for twinning[2] | $< \|L\| > = 0.56$, $< L^2 > = 0.40$ | Xtriage |
| Estimated twinning fraction | 0.000 for h,-k,-l | Xtriage |
| $F_o$,$F_c$ correlation | 0.98 | EDS |
| Total number of atoms | 451 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 15.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 61.56 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.2857e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

# 5 Model quality

## 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 18Y, DMS, GUN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

There are no protein, RNA or DNA chains available to summarize Z scores of covalent bonds and angles.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|---|---|---|---|---|---|---|
| 1 | A | 49 | 41 | 9 | 0 | 0 |
| 1 | B | 47 | 37 | 37 | 0 | 0 |
| 1 | C | 47 | 37 | 38 | 1 | 0 |
| 1 | D | 49 | 37 | 13 | 0 | 0 |
| 2 | A | 11 | 4 | 0 | 0 | 0 |
| 2 | B | 11 | 4 | 5 | 1 | 0 |
| 2 | C | 11 | 4 | 5 | 1 | 0 |
| 2 | D | 11 | 4 | 5 | 0 | 0 |
| 3 | C | 8 | 12 | 12 | 0 | 0 |
| 3 | D | 8 | 12 | 12 | 0 | 0 |
| 4 | D | 7 | 0 | 0 | 0 | 0 |
| All | All | 259 | 192 | 136 | 2 | 0 |

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 11.

All (2) close contacts within the same asymmetric unit are listed below, sorted by their clash

magnitude.

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|---|---|---|---|
| 2:B:102:GUN:HN9 | 2:C:102:GUN:HN21 | 1.40 | 0.68 |
| 1:C:101:18Y:O3' | 1:C:101:18Y:C8' | 2.67 | 0.43 |

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

There are no protein molecules in this entry.

### 5.3.2 Protein sidechains

There are no protein molecules in this entry.

### 5.3.3 RNA

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

16 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the

expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Type | Chain | Res | Link | Bond lengths | | | Bond angles | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Counts | RMSZ | $\#\|Z\|>2$ | Counts | RMSZ | $\#\|Z\|>2$ |
| 3 | DMS | C | 107 | - | 3,3,3 | 0.61 | 0 | 3,3,3 | 1.60 | 0 |
| 1 | 18Y | C | 101 | 2 | 49,53,53 | 2.57 | 19 (38%) | 57,88,88 | 2.52 | 15 (26%) |
| 3 | DMS | D | 108[A] | - | 3,3,3 | 0.68 | 0 | 3,3,3 | 1.07 | 0 |
| 3 | DMS | D | 108[B] | - | 3,3,3 | 0.86 | 0 | 3,3,3 | 1.26 | 0 |
| 2 | GUN | C | 102 | 1 | 7,12,12 | 1.75 | 3 (42%) | 8,17,17 | 1.11 | 1 (12%) |
| 1 | 18Y | A | 101[C] | - | 48,51,53 | 4.27 | 25 (52%) | 57,85,88 | 2.35 | 16 (28%) |
| 2 | GUN | D | 102 | 1 | 7,12,12 | 1.60 | 1 (14%) | 8,17,17 | 1.76 | 3 (37%) |
| 2 | GUN | B | 102 | 1 | 7,12,12 | 1.23 | 1 (14%) | 8,17,17 | 0.96 | 0 |
| 1 | 18Y | B | 101 | 2 | 49,53,53 | 3.04 | 25 (51%) | 57,88,88 | 2.86 | 18 (31%) |
| 1 | 18Y | A | 101[A] | - | 48,51,53 | 4.28 | 26 (54%) | 57,85,88 | 2.43 | 17 (29%) |
| 1 | 18Y | A | 101[B] | - | 48,51,53 | 4.27 | 25 (52%) | 57,85,88 | 2.43 | 16 (28%) |
| 1 | 18Y | D | 101[A] | - | 49,53,53 | 3.34 | 20 (40%) | 57,88,88 | 2.86 | 19 (33%) |
| 1 | 18Y | D | 101[B] | - | 49,53,53 | 3.33 | 20 (40%) | 57,88,88 | 2.83 | 20 (35%) |
| 3 | DMS | C | 106 | - | 3,3,3 | 0.62 | 0 | 3,3,3 | 0.27 | 0 |

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

| Mol | Type | Chain | Res | Link | Chirals | Torsions | Rings |
|---|---|---|---|---|---|---|---|
| 1 | 18Y | C | 101 | 2 | - | 3/23/100/100 | 0/1/7/7 |
| 2 | GUN | C | 102 | 1 | - | - | 0/2/2/2 |
| 1 | 18Y | A | 101[C] | - | - | 4/17/94/100 | 0/1/7/7 |
| 2 | GUN | D | 102 | 1 | - | - | 0/2/2/2 |
| 2 | GUN | B | 102 | 1 | - | - | 0/2/2/2 |
| 1 | 18Y | B | 101 | 2 | - | 4/23/100/100 | 0/1/7/7 |
| 1 | 18Y | A | 101[A] | - | - | 2/17/94/100 | 0/1/7/7 |
| 1 | 18Y | A | 101[B] | - | - | 3/17/94/100 | 0/1/7/7 |
| 1 | 18Y | D | 101[A] | - | - | 5/23/100/100 | 0/1/7/7 |
| 1 | 18Y | D | 101[B] | - | - | 6/23/100/100 | 0/1/7/7 |

The worst 5 of 165 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 1 | A | 101[A] | 18Y | O1'-C5' | -15.44 | 1.23 | 1.44 |
| 1 | A | 101[B] | 18Y | O1'-C5' | -15.44 | 1.23 | 1.44 |
| 1 | A | 101[C] | 18Y | O1'-C5' | -15.44 | 1.23 | 1.44 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|---|---|---|---|---|---|---|---|
| 1 | A | 101[A] | 18Y | O3'-C3' | -11.74 | 1.24 | 1.43 |
| 1 | A | 101[B] | 18Y | O3'-C3' | -11.74 | 1.24 | 1.43 |

The worst 5 of 125 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed($^o$) | Ideal($^o$) |
|---|---|---|---|---|---|---|---|
| 1 | C | 101 | 18Y | C9-C20-C1 | 12.55 | 124.07 | 119.50 |
| 1 | B | 101 | 18Y | C27-O26-C25 | 11.86 | 133.54 | 113.64 |
| 1 | D | 101[A] | 18Y | O16-C25-O26 | -8.98 | 85.95 | 110.02 |
| 1 | A | 101[A] | 18Y | O3'-C3'-C2' | 8.21 | 125.85 | 109.12 |
| 1 | A | 101[B] | 18Y | O3'-C3'-C2' | 8.21 | 125.85 | 109.12 |

There are no chirality outliers.

5 of 27 torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|---|---|---|---|---|
| 1 | A | 101[C] | 18Y | C14-C15-C25-O26 |
| 1 | D | 101[B] | 18Y | O11-C15-C25-O16 |
| 1 | D | 101[B] | 18Y | O12-C15-C25-O16 |
| 1 | D | 101[B] | 18Y | C14-C15-C25-O16 |
| 1 | B | 101 | 18Y | O16-C25-O26-C27 |

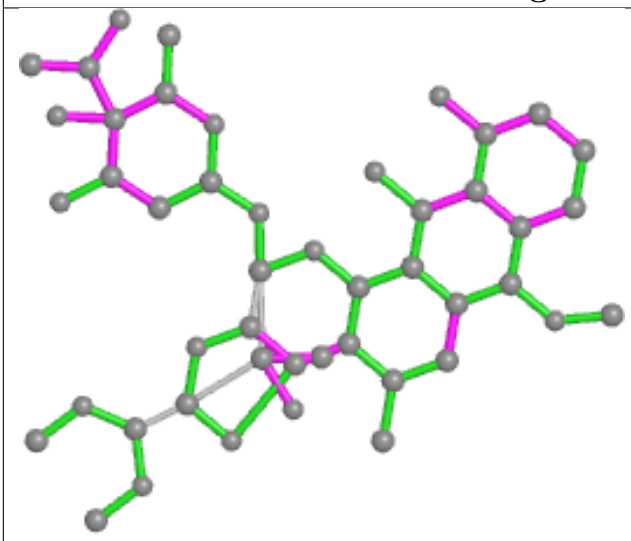
There are no ring outliers.

3 monomers are involved in 2 short contacts:

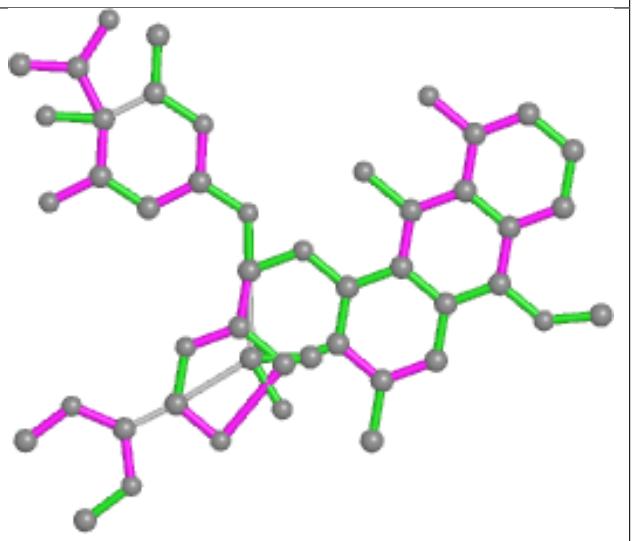
| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|---|---|---|---|---|---|
| 1 | C | 101 | 18Y | 1 | 0 |
| 2 | C | 102 | GUN | 1 | 0 |
| 2 | B | 102 | GUN | 1 | 0 |

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

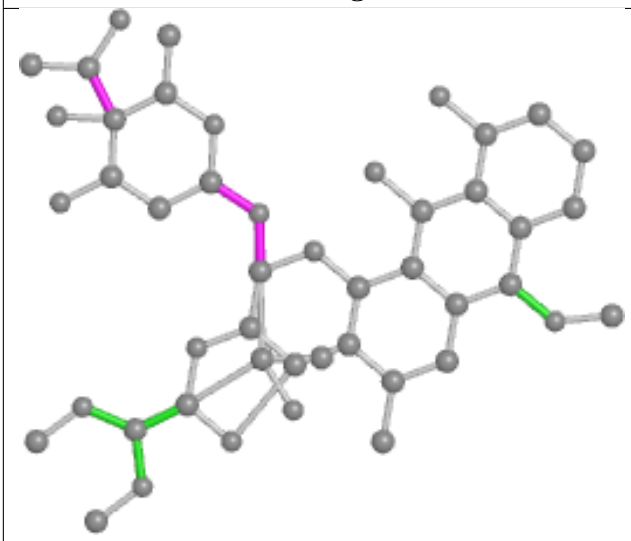
**Ligand 18Y C 101**



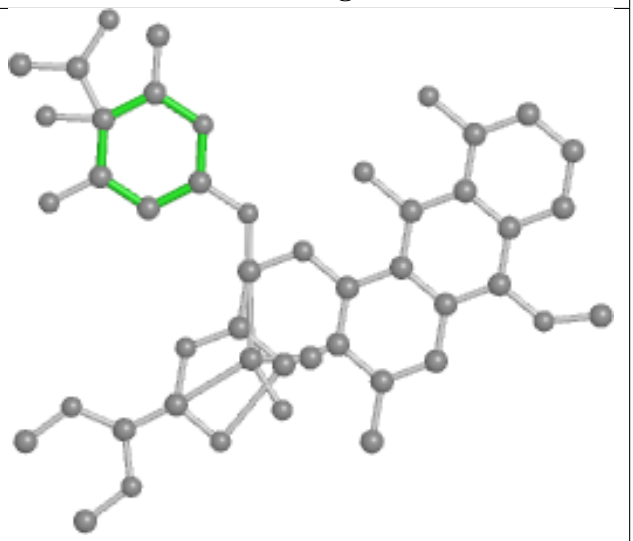
**Bond lengths**



**Bond angles**

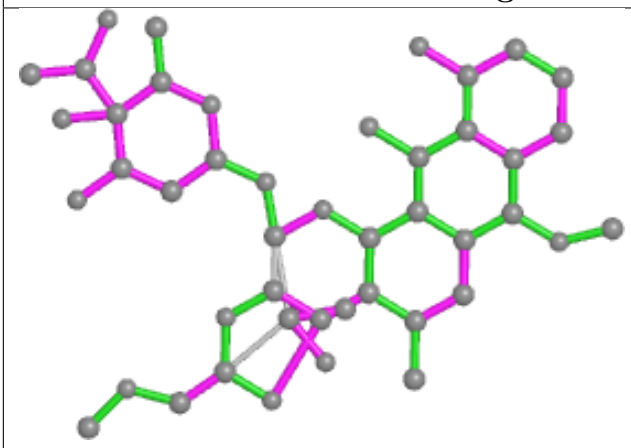


**Torsions**

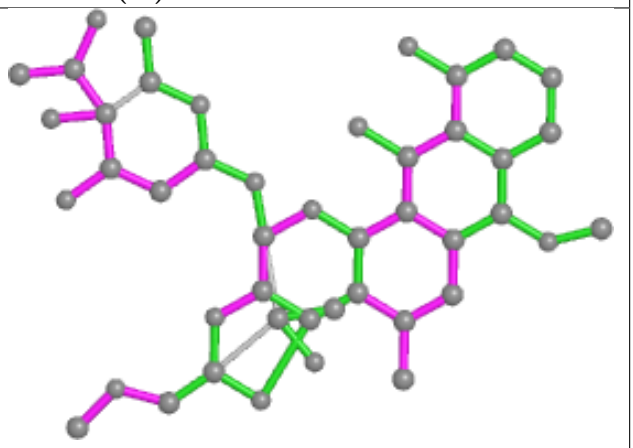


**Rings**

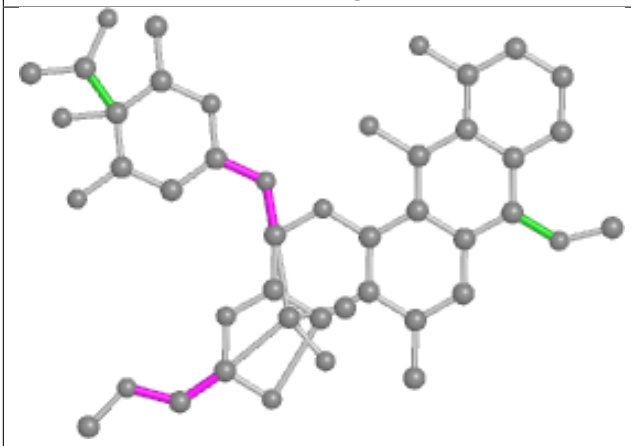
**Ligand 18Y A 101 (C)**



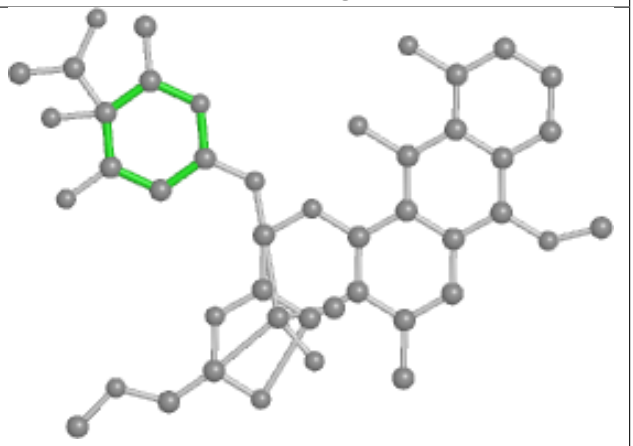
**Bond lengths**



**Bond angles**

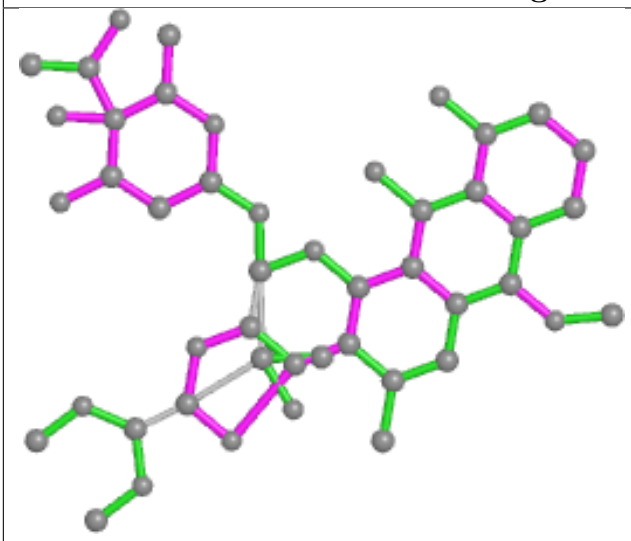


**Torsions**

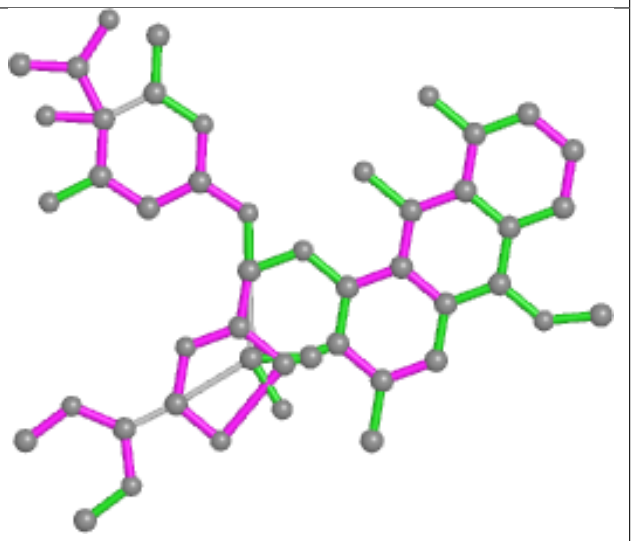


**Rings**

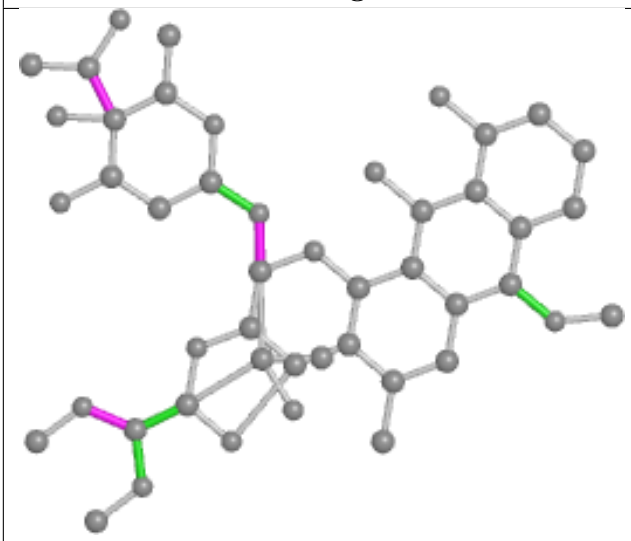
**Ligand 18Y B 101**



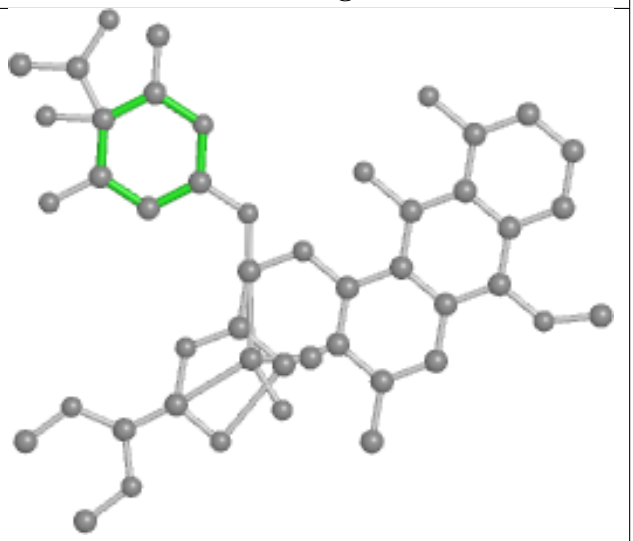
**Bond lengths**



**Bond angles**

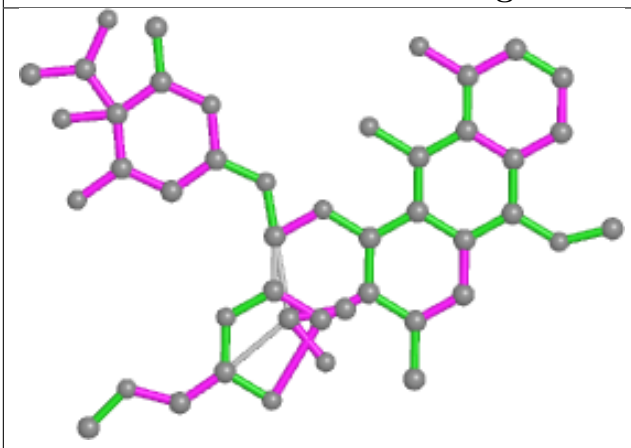


**Torsions**

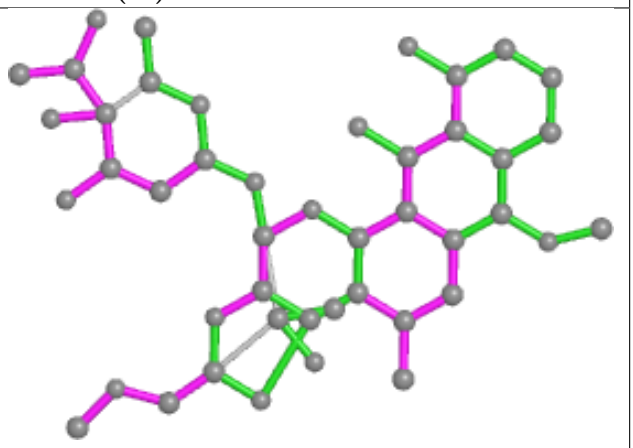


**Rings**

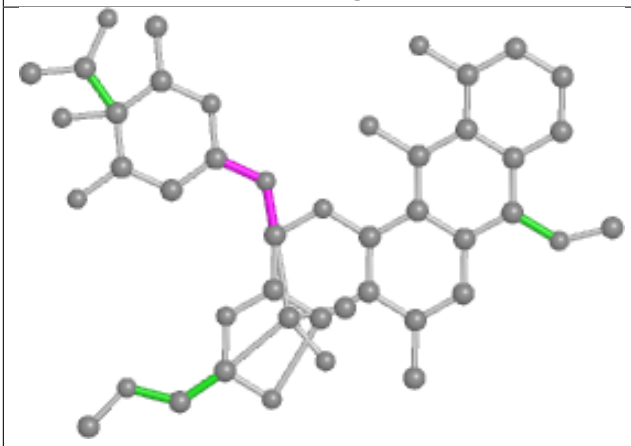
**Ligand 18Y A 101 (A)**



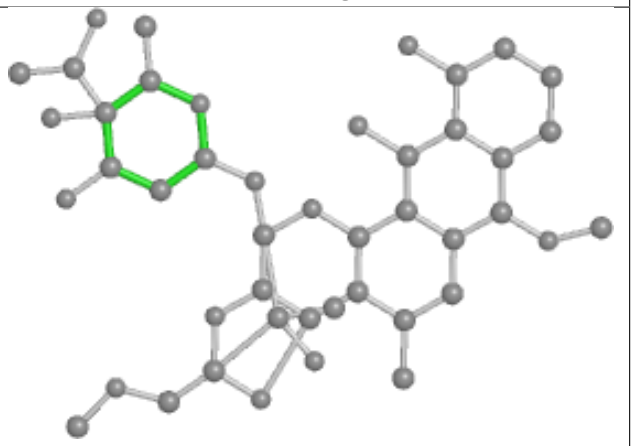
**Bond lengths**



**Bond angles**

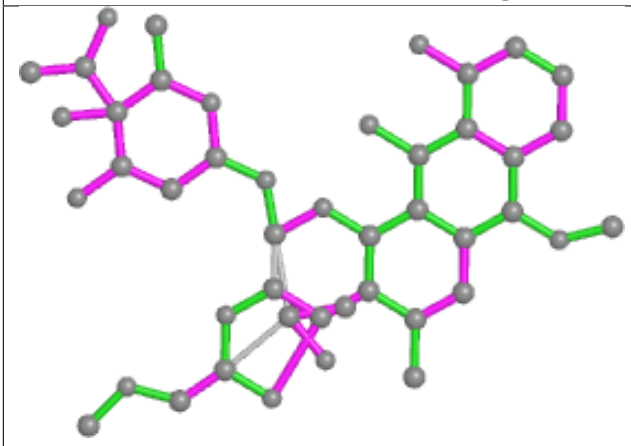


**Torsions**

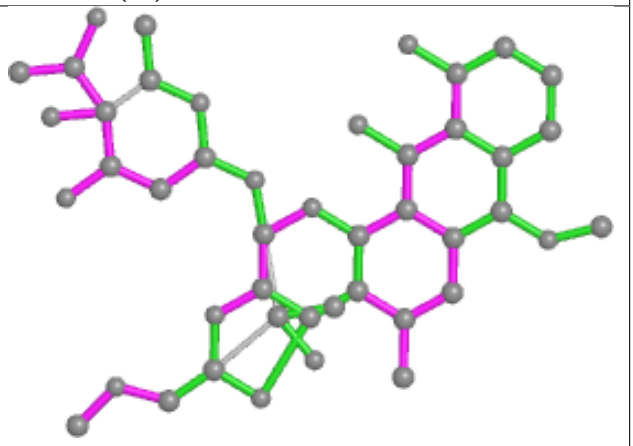


**Rings**

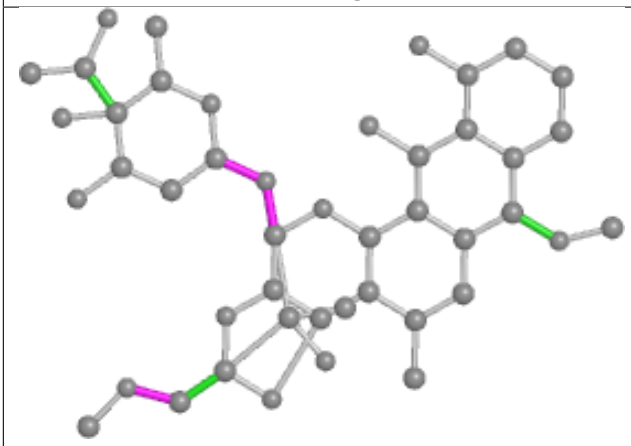
**Ligand 18Y A 101 (B)**



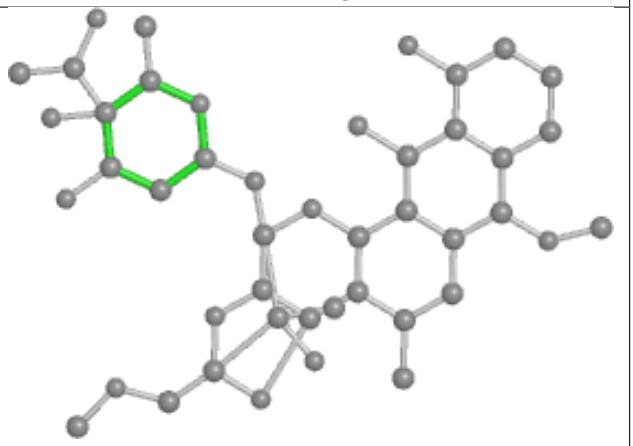
**Bond lengths**



**Bond angles**

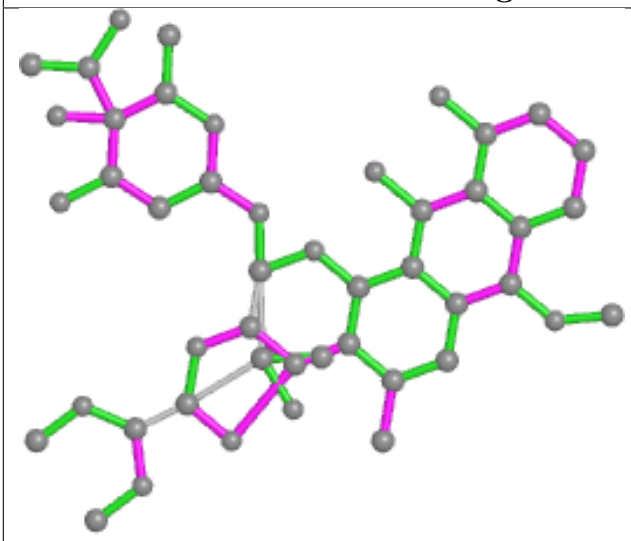


**Torsions**

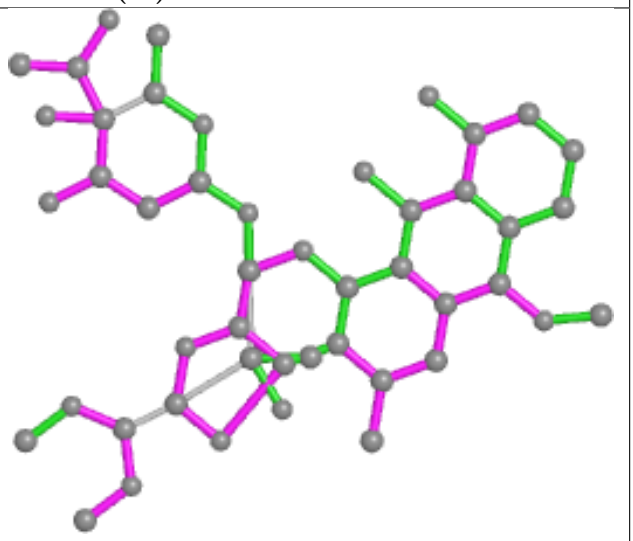


**Rings**

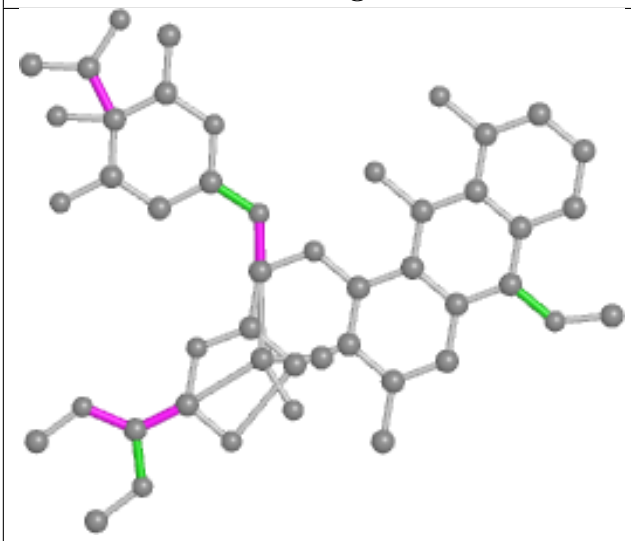
**Ligand 18Y D 101 (A)**



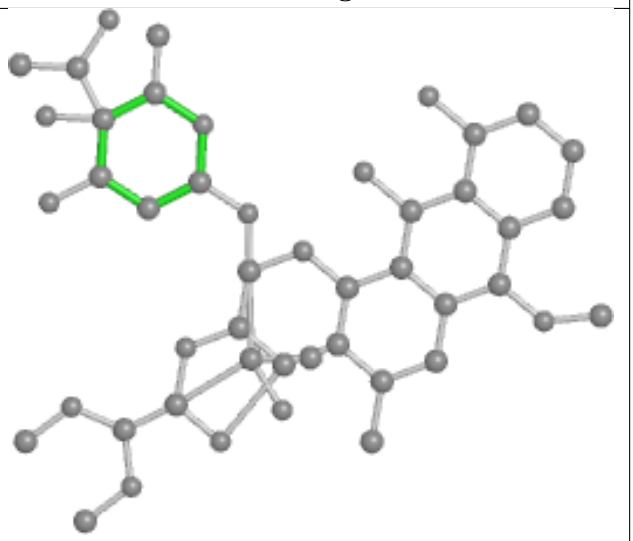
**Bond lengths**



**Bond angles**

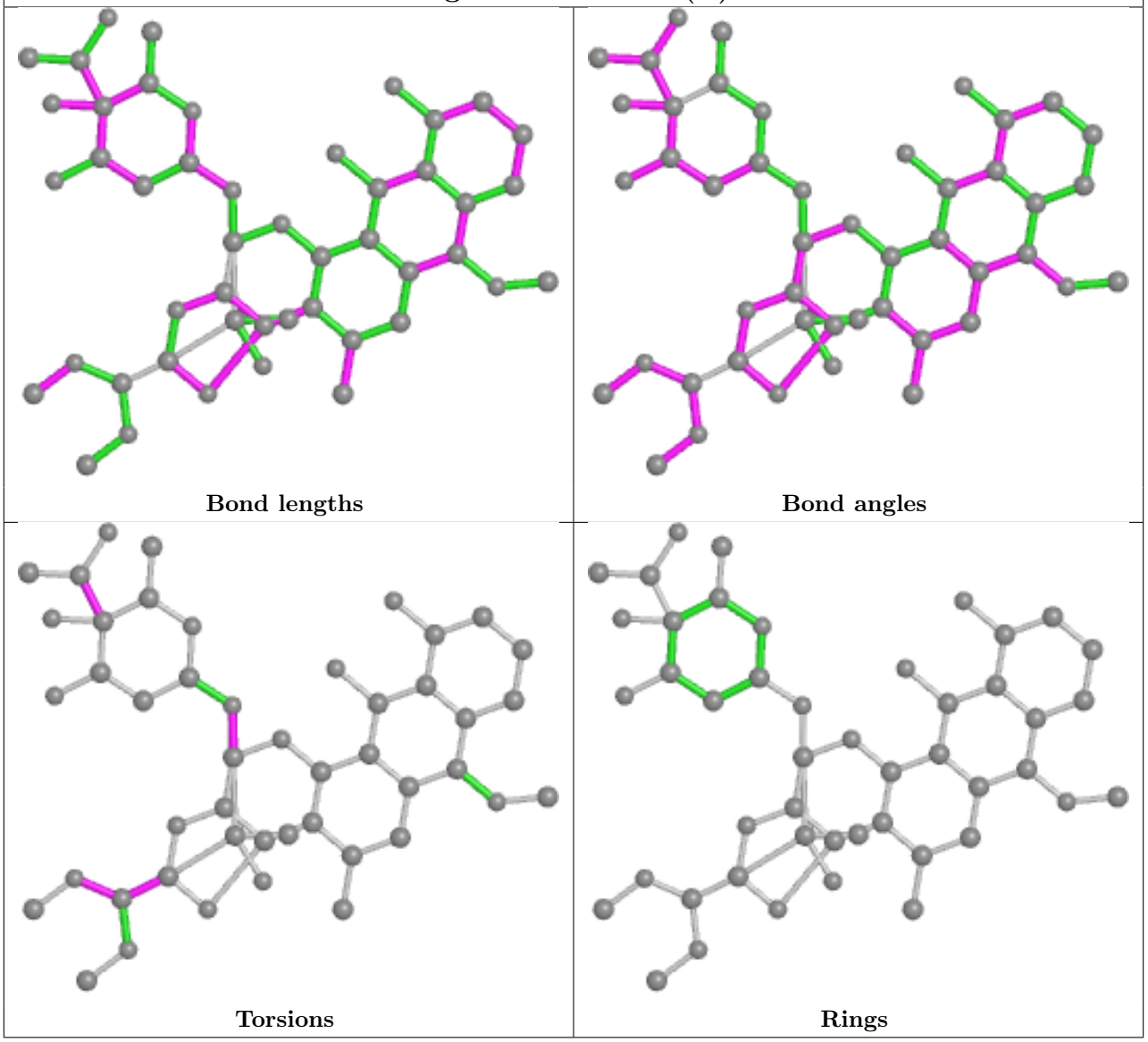


**Torsions**



**Rings**

**Ligand 18Y D 101 (B)**




## 5.7 Other polymers (i)

There are no such residues in this entry.

## 5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

# 6 Fit of model and data

## 6.1 Protein, DNA and RNA chains

There are no RSRZ outliers to report within protein, DNA, RNA chains in this entry.

## 6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates

There are no monosaccharides in this entry.

## 6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, $95^{th}$ percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

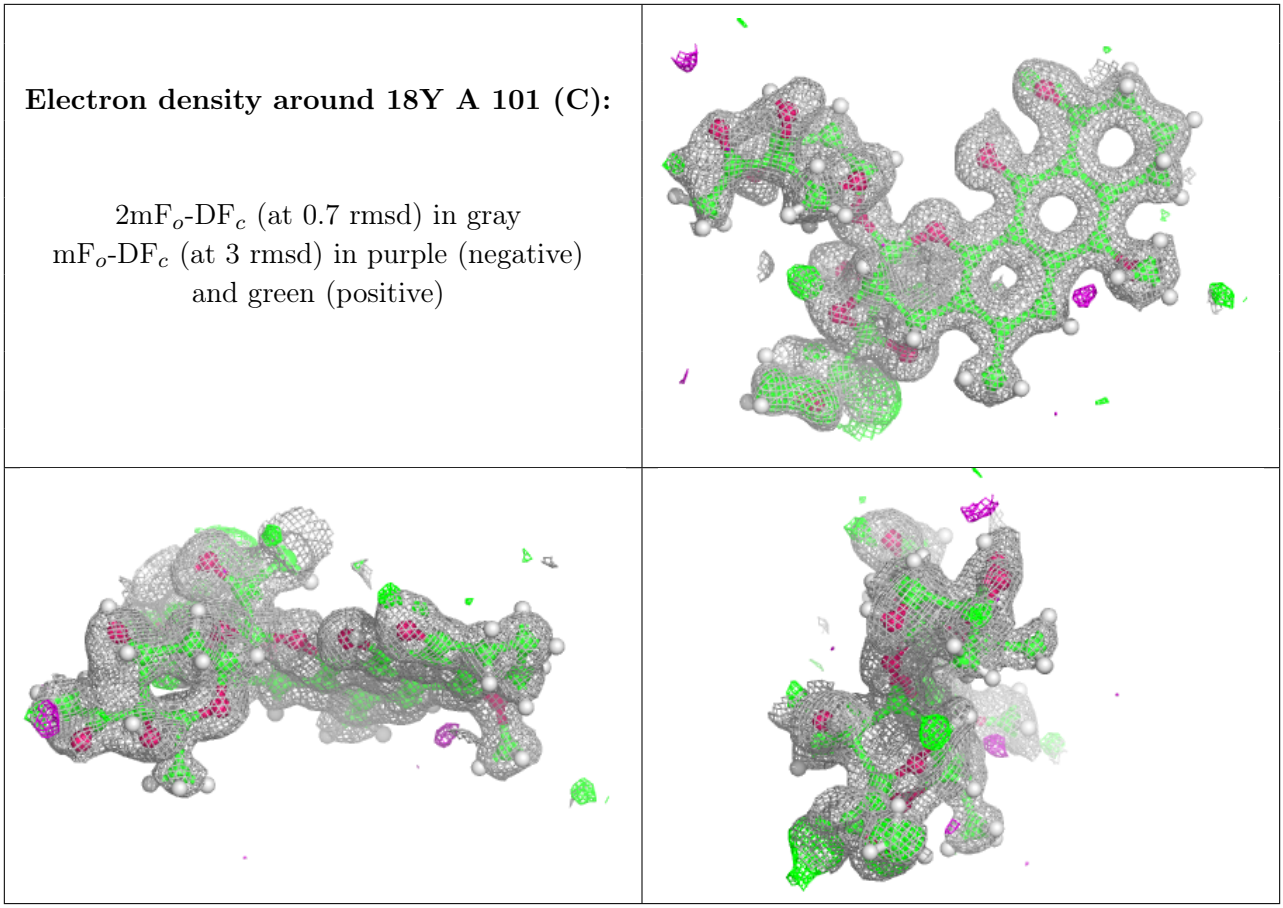
| Mol | Type | Chain | Res | Atoms | RSCC | RSR | B-factors(Å$^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 3 | DMS | D | 108[A] | 4/4 | 0.92 | 0.39 | 15,22,23,23 | 10 |
| 3 | DMS | D | 108[B] | 4/4 | 0.92 | 0.39 | 15,22,23,23 | 10 |
| 1 | 18Y | A | 101[C] | 45/47 | 0.94 | 0.10 | 5,10,34,51 | 5 |
| 3 | DMS | C | 106 | 4/4 | 0.94 | 0.17 | 15,22,23,23 | 0 |
| 1 | 18Y | A | 101[A] | 45/47 | 0.94 | 0.10 | 5,10,51,52 | 5 |
| 1 | 18Y | A | 101[B] | 45/47 | 0.94 | 0.10 | 5,10,51,52 | 5 |
| 1 | 18Y | D | 101[B] | 47/47 | 0.95 | 0.09 | 6,11,31,34 | 11 |
| 1 | 18Y | D | 101[A] | 47/47 | 0.95 | 0.09 | 6,11,31,34 | 11 |
| 1 | 18Y | C | 101 | 47/47 | 0.96 | 0.08 | 6,11,28,31 | 0 |
| 1 | 18Y | B | 101 | 47/47 | 0.96 | 0.08 | 6,13,31,47 | 0 |
| 3 | DMS | C | 107 | 4/4 | 0.97 | 0.09 | 15,23,29,29 | 0 |
| 2 | GUN | A | 102[B] | 11/11 | 0.97 | 0.08 | 5,6,7,8 | 1 |
| 2 | GUN | A | 102[A] | 11/11 | 0.97 | 0.08 | 5,6,7,8 | 1 |
| 2 | GUN | C | 102 | 11/11 | 0.98 | 0.06 | 4,6,8,9 | 0 |
| 2 | GUN | D | 102 | 11/11 | 0.98 | 0.07 | 5,6,7,8 | 0 |
| 2 | GUN | B | 102 | 11/11 | 0.98 | 0.07 | 4,5,6,7 | 0 |

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different

orientation to approximate a three-dimensional view.

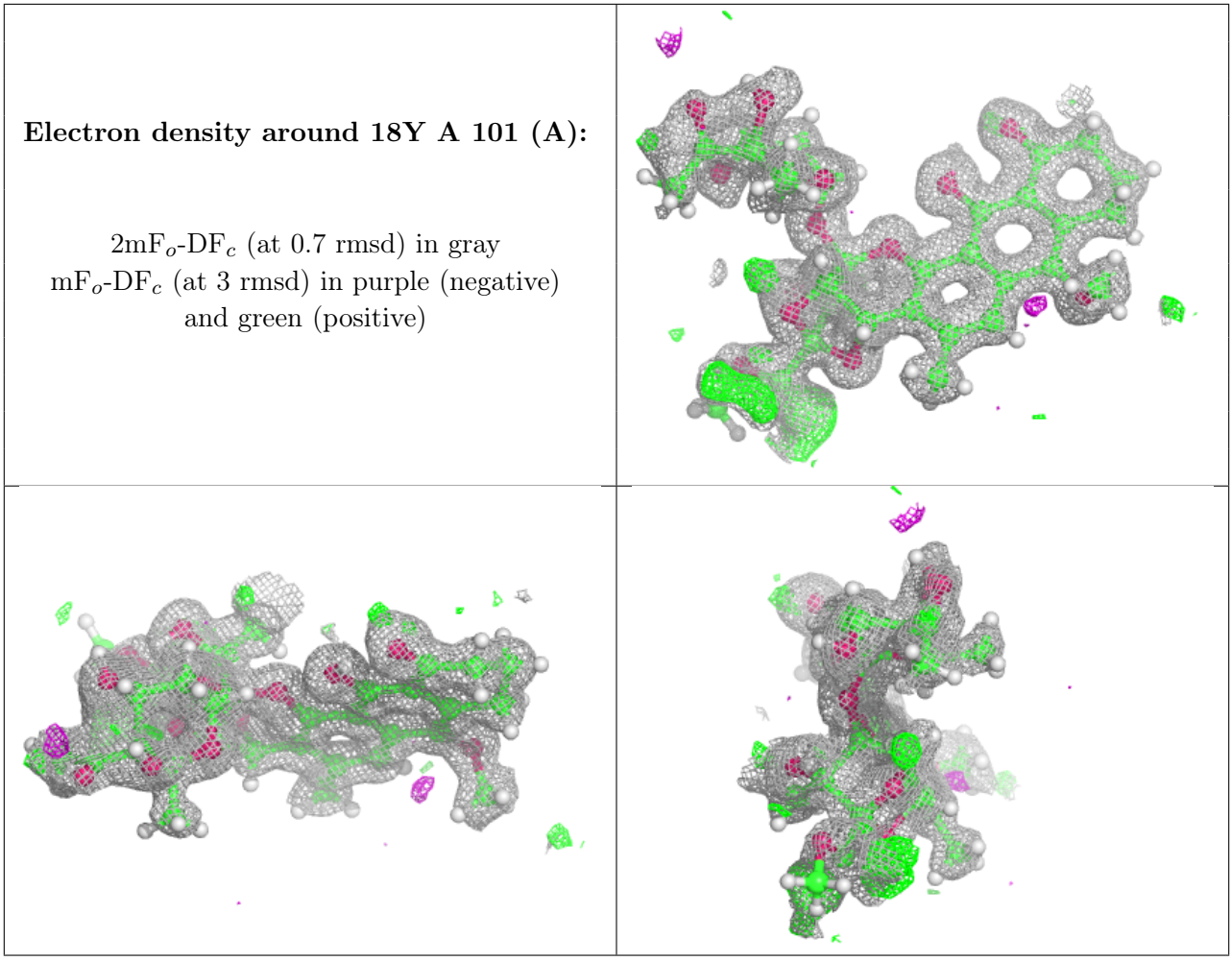
**Electron density around 18Y A 101 (C):**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



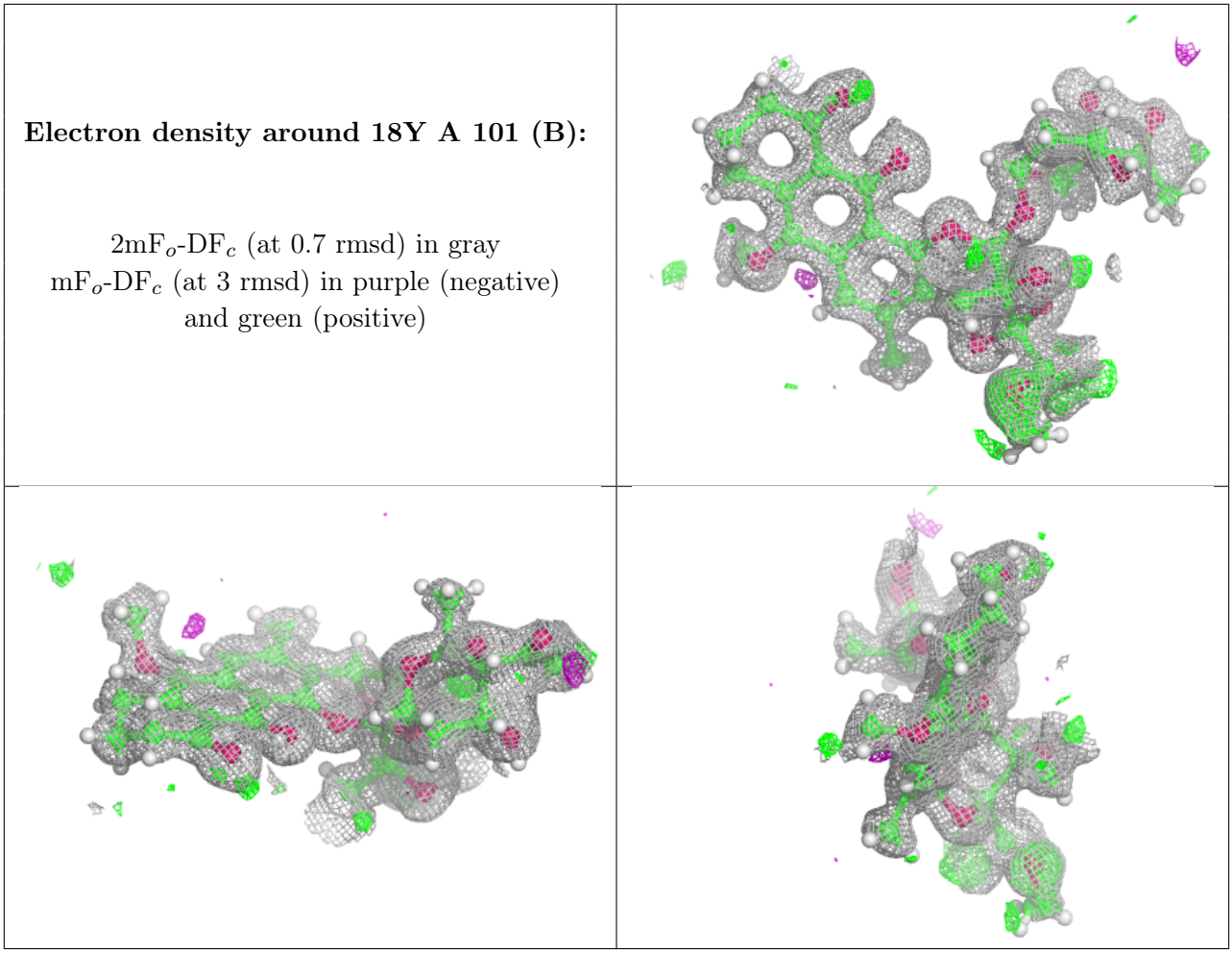
**Electron density around 18Y A 101 (A):**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



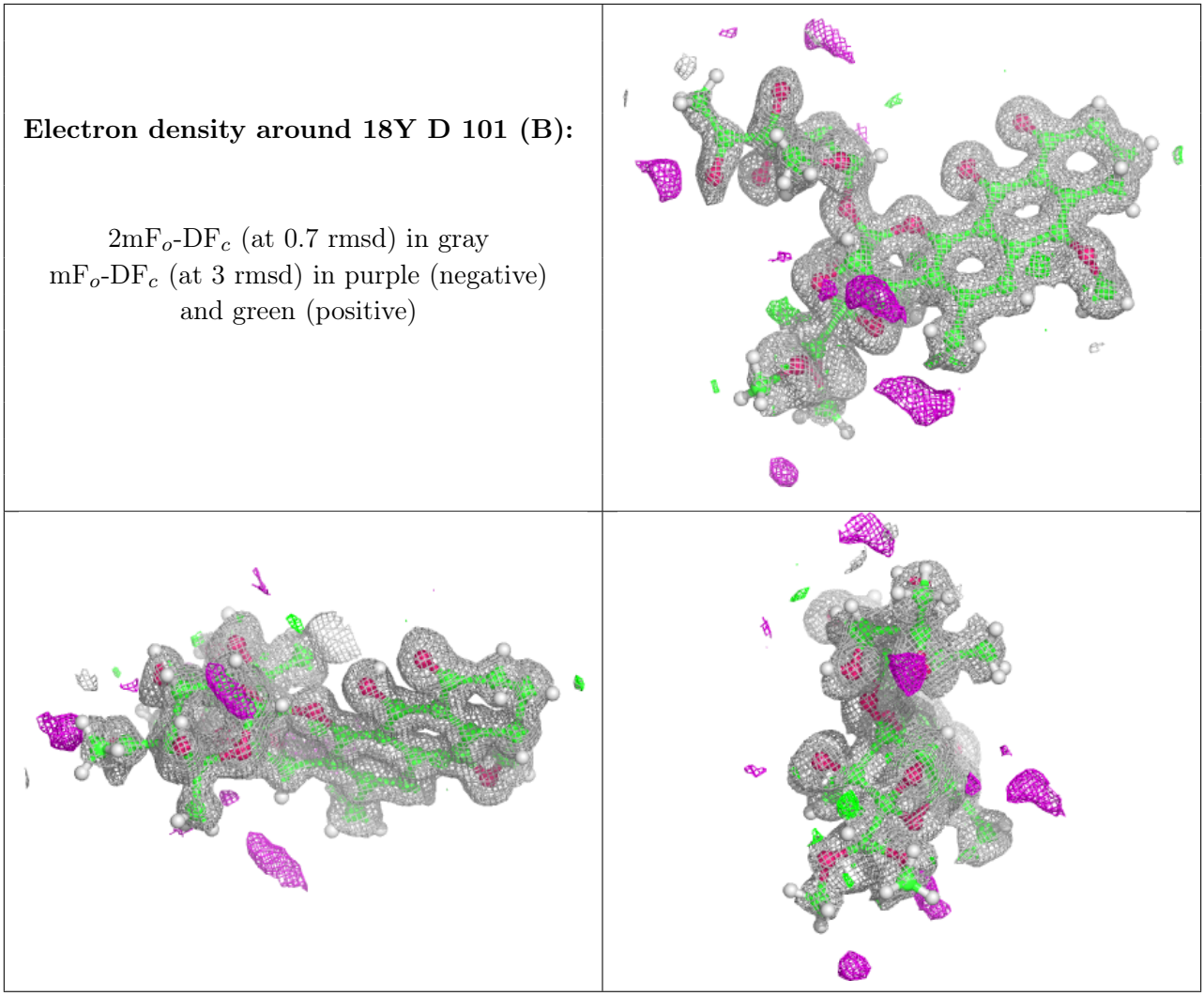
**Electron density around 18Y A 101 (B):**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



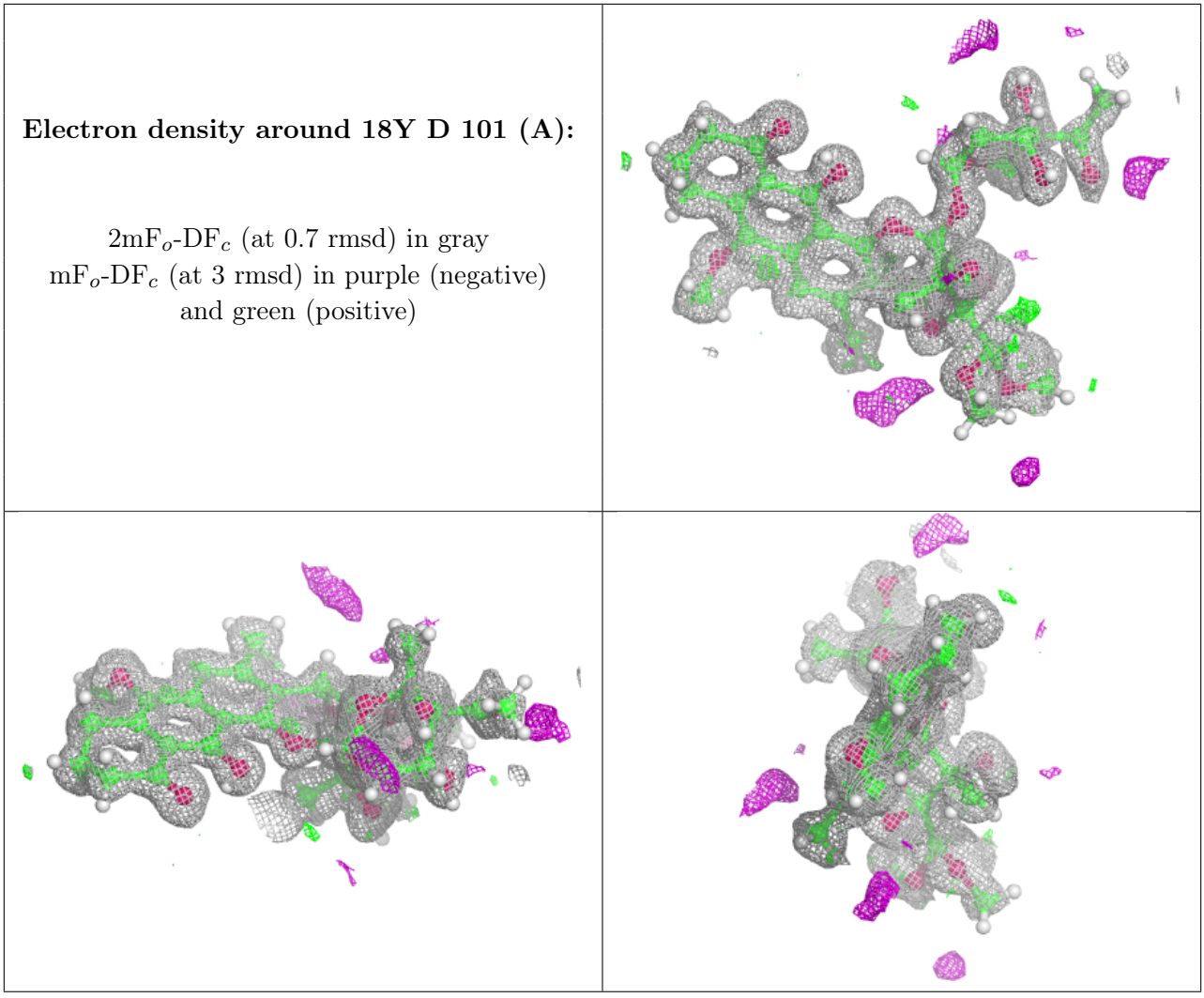
**Electron density around 18Y D 101 (B):**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



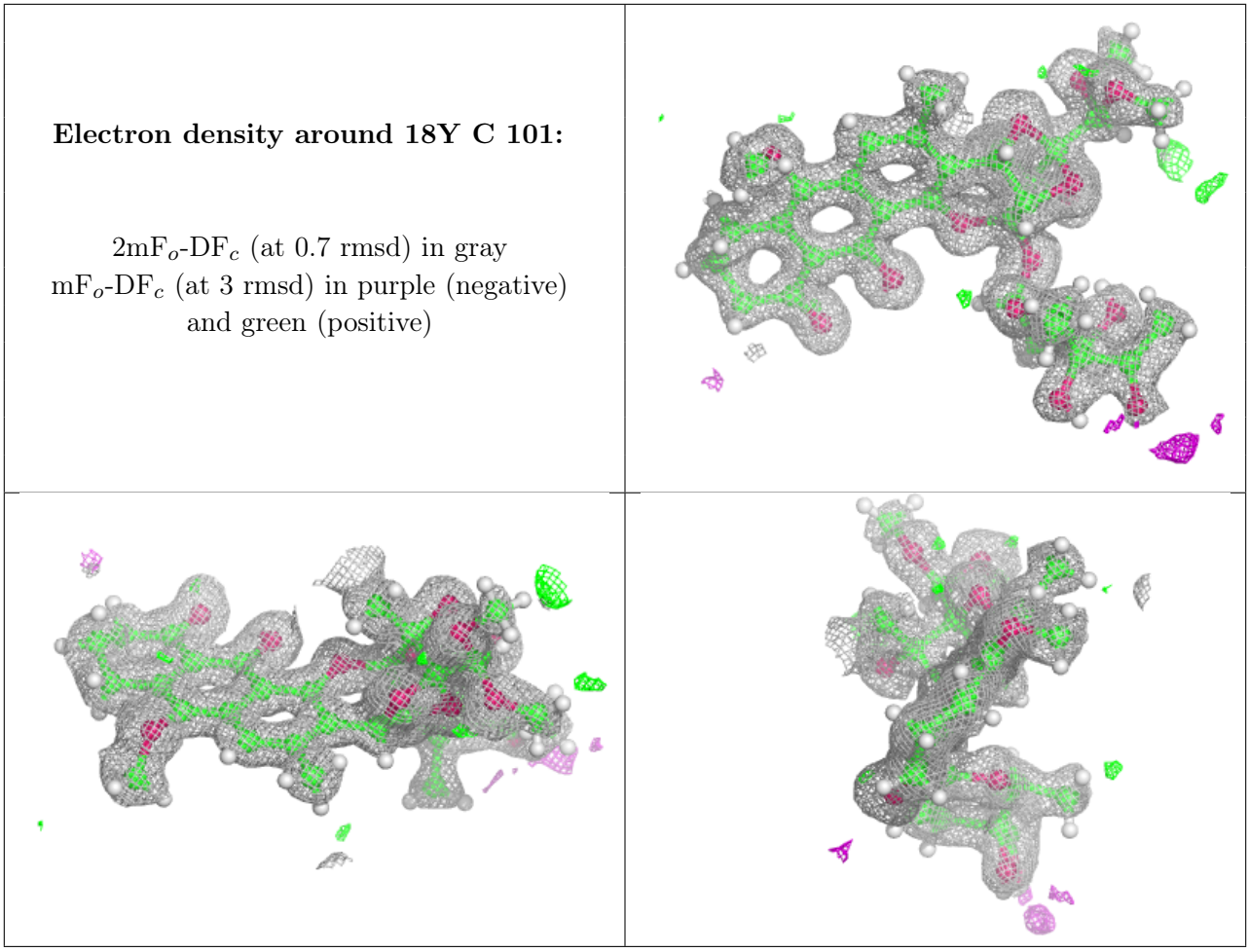
**Electron density around 18Y D 101 (A):**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



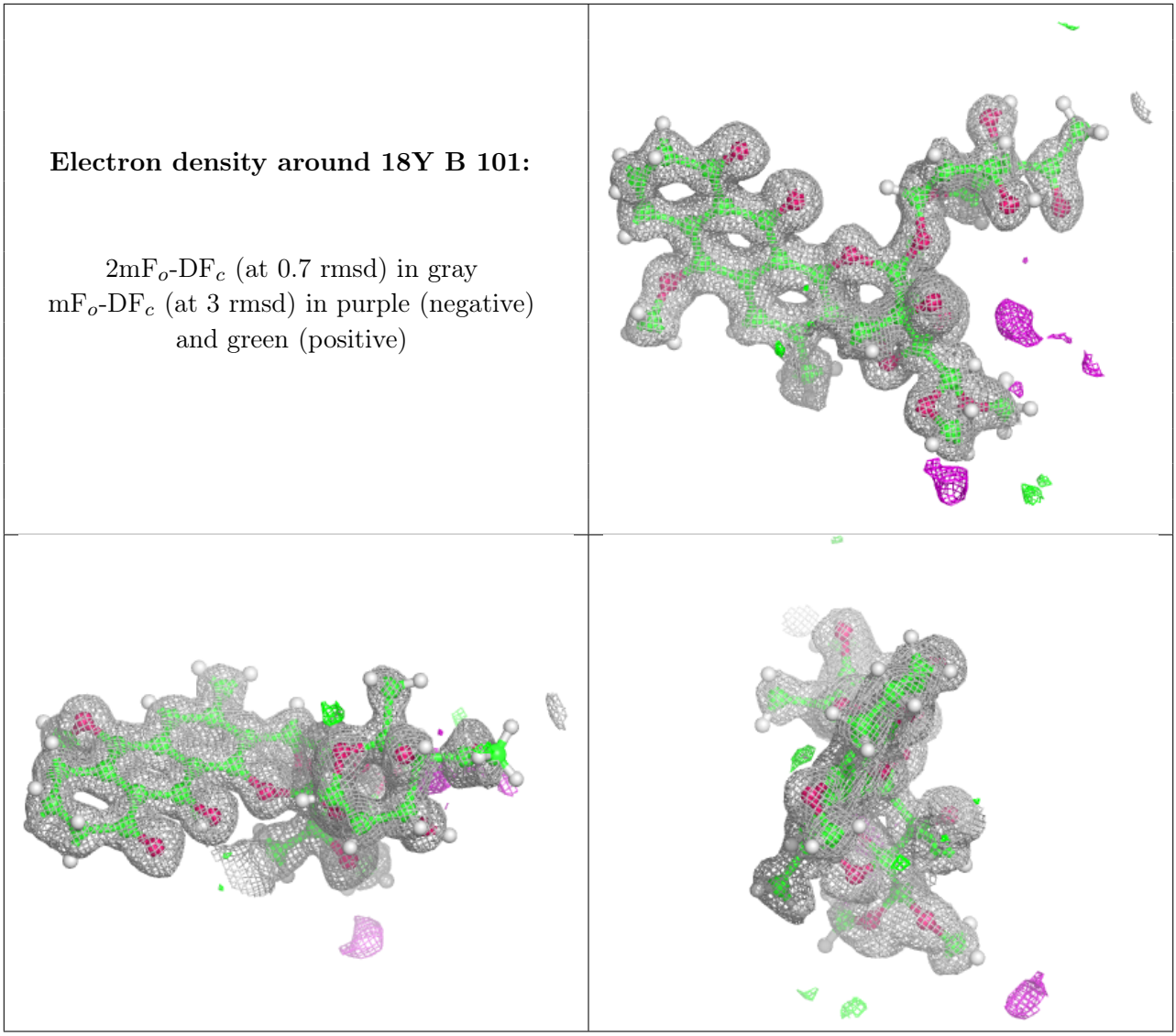
**Electron density around 18Y C 101:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



**Electron density around 18Y B 101:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative) and green (positive)



## 6.5 Other polymers (i)

There are no such residues in this entry.